

# Digitale Archive

Johannes H. Stigler

Zentrum für Informationsmodellierung in den Geisteswissenschaften  
Universität Graz

Sommersemester 2017



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



# Organisatorisches

## Erreichbarkeit

- johannes.stigler@uni-graz.at
- 0316 380 2292
- Sprechstunde nach Vereinbarung

## Zur Lehrveranstaltung

- Termine und Unterlagen zur Lehrveranstaltung finden sich auf der Lernplattform Moodle (<http://moodle.uni-graz.at/>)
- Benotung: Mitarbeit, Hausübungen, ev. abschließendes Prüfungsgespräch
- Anwesenheit: prüfungsimmanenter Charakter



# Agenda

- 1 Organisatorisches
- 2 **Software**
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



# Software

- Java Runtime Environment <https://www.java.com/de/>
- Cirilo Administration Client <http://gams.uni-graz.at/cirilo/2.0>
- Oxygen XML-Editor <http://www.oxygenxml.com/download.html>



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 **Allgemeines**
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



# What is Digital Preservation?

- Sicherstellung der authentischen und nachhaltigen Verfügbarkeit digitaler Ressourcen
- Zunehmend wird digital produziert, verbreitet, genutzt („born digital“, Aufsätze, Working Papers, Conference Proceedings, Projektberichte u.a.)
- Open Access als Erfordernis der Forschungsförderung
- LZA als Digitalisierung analoger Ressourcen, als Sicherung des „digitalen Erbes“ für die Zukunft
- Archivierung ist nicht nur eine Frage von technischen Lösungen sondern erfordert auch institutionelle Strategien und Policies
- Archivierung beginnt jedenfalls mit der Datenproduktion und erfordert Überlegungen zur Standardisierungen von Datenformaten
- Aus Daten werden Information über die Erschließung des Archivierungsgegenstandes mit Metadaten
- Archivierung dient der Bestandserhaltung und der Dissemination





# What is Digital Preservation?

Beschäftigung mit Fragen der Nachhaltigkeit in einer kurzlebigen IT-Welt

## Problemstellung

- Erhaltung des Binärstroms (z.B. LOCKSS Lots Of Copies Keep Stuff Safe)
- Erhaltung der Interpretierbarkeit (Daten- und Metadatenformate wie PDF/A, TEI)

## Lösungen

- Spezielle Archivsysteme
- Maßnahmen (Migration, Emulation)



# Repository Service Levels

## Bitstream-Preservation

- Redundante Offline-Speicherung
- für 10 Jahre

## Einfache LZA

- Level 1+
- Just-in-time-Zugriff

## Content-Preservation

- Level 2+
- Technisches Qualitätsmanagement

## Data Curation

- Level 3+
- Kontinuierliche Pflege der Metadaten und inhaltlichen Bezüge
- Dissemination der Daten und Metadaten über Webschnittstelle

Tabelle: nach Pempe 2012



# What is a Digital Archive?

## Allgemein

- Aus inhaltlicher Perspektive: Eine Sammlung digital repräsentierter, erschlossener kultureller Artefakte (Texte, Bilder u.a.)
- Aus archivischer Perspektive: Eine organisatorische und technologische Infrastruktur zur Sicherstellung von Nachhaltigkeit unter der Prämisse einer sich ständig wandelnden IT-Technologie
- Digitale Archive dienen der Archivierung, Verwaltung und Bereitstellung digitaler Ressourcen
- Der „Idelatypus“ OAIS (Open Archival Information System) Referenzmodell



# What is a Digital Archive?

## Funktionale Anforderungen

- Direkter Zugriff auf einzelne Informationsobjekte oder Informationskollektionen mittels Persistent Identifier
- Einheitliche und gemeinsame Speicherung beliebiger Informationsobjekte vom gesamten Facsimilies, über XML-Strukturen und Video- bzw. Audiostreams
- Konverter zur Erzeugung von langfristig stabilen Archivformaten. Wenn nötig transformiert es Inhalte in nachhaltige, technologieunabhängige (für die Langzeitarchivierung geeignete) Datenformate
- Dokumentation aller Veränderungen an Strukturen und Informationsobjekten: Das Digitale Archiv besitzt „Wissen“ über Inhalt und Lebenszyklen aller Objekte
- Informationsobjekte werden über eine Rechteverwaltung gegen unberechtigten Zugriff und Veränderung der gespeicherten Inhalte geschützt
- Bei Bedarf sorgt es für automatisierte, nachvollziehbare und verlustfreie Migrationsverfahren und für die Dokumentation aller Veränderungen an Strukturen und Informationsobjekten im Archiv



# What is a Digital Archive?

- Organisierte Sammlung digitaler Objekte bestehend aus z.B. Text-, Bild-, Ton-, Video- und Multimediadatenströmen
- Die (digitalen) Objekte werden inhaltlich (z.B. TEI) und bibliographisch (z.B. Dublin Core) nach Standards erschlossen (Metadaten)
- Die digitalen Objekte werden über Interfaces, Services und Schnittstellen (z.B. OAI-PMH) nachhaltig zugänglich gemacht
- Die digitalen Objekte erhalten eine eindeutige, persistente und zitable Identifizierung (z.B. DOI, URN, PURL)
- Die digitalen Objekte werden unter Sicherstellung ihrer Authentizität (Prüfsummen, digitale Signaturen etc.) langfristig bewahrt



# What is a Digital Archive?

## Dokumentenserver

Publikationsplattformen unter der Prämisse des Open Access. Das DINI-Zertifikat für Dokumentenserver zielt auf folgende Aspekte ab:

- Leitlinien (Policy)
- Autorenbetreuung
- Rechtliche Aspekte
- Authentizität und Integrität
- Erschließung
- Interoperabilität
- Zugriffsstatistik
- Langzeitverfügbarkeit

<http://www.dini.de/documents/Zertifikat.pdf>



# What is a Digital Archive?

## Information objects ...

## VASE

### Visual Archive Southeastern Europe

Home About Terms of Use Photographers & Artists Literature Partner Institutions

Advanced Search

#### Šeher-Čehajin Bridge and Bendula

**Description**

**Object:** Šeher-Čehajin Bridge and Bendula

**Description:** People passing the Šeher-Čehajin Bridge: two veiled women and a Muslim woman's street clothes. A man in traditional clothes is near the bridge. A group of men in suits is standing at the bridge head square. In the background: the residence (konak) at Bendula.

**Comment:** The residence (konak) at Bendula was built in 1836 by Mustafa Pasha, known Sarajevo feudal lord.

**Relations:** <http://www.vase.graz.ac.at/vase/2100>  
<http://www.vase.graz.ac.at/vase/2101>  
<http://www.vase.graz.ac.at/vase/2102>

**Date:** Not before 1914, Not after 1918

**Location:** [Bosnia](#)

**Country:** Bosnia and Herzegovina

**Type:** Postcard

**Creator:** [unknown](#)

**Publisher:** Leon Finci, Sarajevo

**Dimensions:** Artifact: 90mm x 135mm

**Format:** Not specified

**Technique:** Not specified

**Keywords:** [200 Animal Subjects](#) • [200 Domesticated Animals](#)  
[200 Clothing](#)  
[200 Landmarks](#)  
[200 People](#)  
[200 Social Institutions](#)  
[200 Social Stratification](#) • [200 Gender Roles](#)  
[200 Religious Practices](#)

**Copyright:** Bošnjački Institut - Fondacija Adila Zulfikarpašića

**Archive:** [Bosnian Institute - Adila Zulfikarpašića Foundation](#), Inv. No.: Not in

**License:** This picture is licensed under [Creative Commons CC BY-NC-ND](#)

**Editor:** Barbara Deier

**Permalink:** <http://www.vase.graz.ac.at/vase/2103>

Content | Search | Register | About | Contact | Privacy Policy | Terms of Use | Copyright |  
Center for Information Literacy in the 21st Century  
University of Graz

GAMS SEEHA southeast european history and anthropology



## VASE

Visual Archive Southeastern Europe

Permalink: <http://www.vase.graz.ac.at/vase/2103>

### Šeher-Čehajin Bridge and Bendula



Šeher-Čehajin Bridge and Bendula  
• Bosnia Institute - Adila Zulfikarpašića Foundation



Šeher-Čehajin Bridge and Bendula  
• Bosnia Institute - Adila Zulfikarpašića Foundation

**Object:** Šeher-Čehajin Bridge and Bendula

**Description:** People passing the Šeher-Čehajin Bridge: two veiled women and two girls are wearing Muslim women's street clothes. A man in traditional clothes is leading a carriage across the bridge. A group of men in suits is standing at the bridge head at Mustafa Pasha's square. In the background: the residence (konak) at Bendula.

**Comment:** The residence (konak) at Bendula was built in 1836 by Mustafa Pasha. He was a well-known Sarajevo feudal lord.

**Relations:** <http://www.vase.graz.ac.at/vase/2100>  
<http://www.vase.graz.ac.at/vase/2101>  
<http://www.vase.graz.ac.at/vase/2102>

**Date:** Not before 1914, Not after 1918

**Location:** Sarajevo

**Country:** Bosnia and Herzegovina

**Type:** Postcard

**Creator:** unknown

**Publisher:** Leon Finci, Sarajevo

**Dimensions:** Artifact: 90mm x 135mm

**Format:** Not specified

**Technique:** Not specified

**Keywords:** [200 Animal Subjects](#) • [200 Domesticated Animals](#)  
[200 Clothing](#)  
[200 Landmarks](#)  
[200 People](#)  
[200 Social Institutions](#) • [200 Gender Roles](#)  
[200 Religious Practices](#)

**Copyright:** Bošnjački Institut - Fondacija Adila Zulfikarpašića

**Archive:** [Bosnian Institute - Adila Zulfikarpašića Foundation](#), Inv. No.: Not in

**License:** This picture is licensed under [Creative Commons CC BY-NC-ND 3.0](#) (<http://creativecommons.org/licenses/by-nc-nd/3.0/>)

**Editor:** Barbara Deier

<http://www.vase.graz.ac.at/vase>

# What is a Digital Archive?

## Information objects ...

- Shows the object content in different views
  - <http://gams.uni-graz.at/o:mws-115-879>
- Shows the object content in different views
  - <http://gams.uni-graz.at/o:mws-115-879/sdef:TEI/get>
- Shows the object content in different views
  - <http://gams.uni-graz.at/o:mws-115-879/sdef:TEI/getPDF>
- Shows the datastreams list for the object
  - <http://gams.uni-graz.at/archive/objects/o:mws-115-879/datastreams>
- Shows the methods list for the object
  - <http://gams.uni-graz.at/archive/objects/o:mws-115-879/methods>





# What is a Digital Archive?

## Persistent Identifier

- A persistent identifier (PI) is a long-lasting reference to a digital object
- Persistent identifier conventions and systems:
  - Handle Systems
  - Digital Object Identifiers (DOIs)
  - Archival Resource Keys (ARKs)
  - Persistent Uniform Resource Locators (PURLs)
  - Uniform Resource Names (URNs)
  - Extensible Resource Identifiers (XRI)



# What is a Digital Archive?

## Handle System

- The Handle System is a technology specification for assigning, managing, and resolving persistent identifiers for digital objects
- The system is designed to be scalable to very large numbers of entities
- The Handle System web site provides a series of implementation tools
  - Implementation of the Handle System consists of Local Handle Services that store specific handles
  - The Global Handle Registry is a unique Local Handle Service which stores information on the prefixes (also known as naming authorities) within the Handle System



# What is a Digital Archive?

## Handle System

An id string ...

- is not based on any changeable attributes of the entity (location, ownership, or any other attribute that may change without changing the reference identity)
- is opaque (preferably a plain number: a well known pattern invites assumptions that may be misleading, and meaningful semantics may not translate across languages and may cause trademark conflicts)
- is unique within the system (to avoid collisions and referential uncertainty)
- `http://hdl.handle.net/handle_prefix/identifier_string`
- `hdl:handle_prefix/identifier_string` (with installed browser plugin)
- e.g. `http://hdl.handle.net/11471/505.20.779`
- TinyURL `https://tinyurl.com/zdqpumd`



# What is a Digital Archive?

## Digital Object Identifier

- DOI („Bezeichner für digitale Objekte“) ist ein eindeutiger und dauerhafter Identifikator für digitale Objekte.
- Definiert ein System, das ein Objekt selbst identifiziert und nicht einen Ort, unter dem es momentan abgelegt ist
- Ist vereinfacht ausgedrückt mit ISBN und ISSN vergleichbar
- Um den Zugriff auf identifizierte Objekte zu ermöglichen, stellt die International DOI Foundation (IDF) als Betreiber des DOI-Systems ein System zur Verfügung, das dem DOI den aktuellen Standort eines Dokuments zuordnet.
- Basiert auf dem Open Source Projekt „Handle“.
- DOIs beginnen immer mit 10. und haben die Form 10.ORGANISATION/ID wobei Organisationen (vor allem Verlage) jeweils eine eigene Nummer zugewiesen bekommen und die ID selbst vergeben können.
- Der DOI 10.1000/182 identifiziert das Dokument 182 bei der Organisation 1000 (der DOI-Stiftung selbst), URL <http://dx.doi.org/10.1000/182>
- Als Uniform Resource Identifier werden DOIs die Schema-Kennzeichnung doi: vorangestellt, so dass sie die Form doi:10.ORGANISATION/ID haben



# What is a Digital Archive?

## Standardized protocols for data exchange and metadata harvesting

- The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability
- Data Providers are repositories that expose structured metadata via OAI-PMH
- Service Providers then make OAI-PMH service request to harvest that metadata
- OAI-PMH is a set of six verbs or services that are invoked within HTTP
  - Archival Metadata
    - <http://gams.uni-graz.at/oaiprotocol?verb=Identify>
    - <http://gams.uni-graz.at/oaiprotocol?verb=ListMetadataFormats>
  - Harvesting verbs
    - [http://gams.uni-graz.at/oaiprotocol/?verb=GetRecord&metadataPrefix=oai\\_europeana&identifier=hdl:11471/505.20.779](http://gams.uni-graz.at/oaiprotocol/?verb=GetRecord&metadataPrefix=oai_europeana&identifier=hdl:11471/505.20.779)
    - [http://gams.uni-graz.at/oaiprotocol?verb=ListIdentifiers&metadataPrefix=oai\\_europeana](http://gams.uni-graz.at/oaiprotocol?verb=ListIdentifiers&metadataPrefix=oai_europeana)
    - [http://gams.uni-graz.at/oaiprotocol?verb=ListRecords&metadataPrefix=oai\\_europeana](http://gams.uni-graz.at/oaiprotocol?verb=ListRecords&metadataPrefix=oai_europeana)
    - <http://gams.uni-graz.at/oaiprotocol?verb=ListSets>



# Trusted Digital Archives

Die Research Libraries Group (RLG) nennt als wichtigste Eigenschaften von *Trusted Digital Repositories*:

- Gesicherte organisatorische Struktur und rechtlicher Status
- Finanzielle Nachhaltigkeit
- Technologische und prozedurale Angemessenheit
- Gewährleistung von Daten- und Systemsicherheit
- Dokumentation und Transparenz
- Konformität mit dem OAIS-Standard

<http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>



# Trusted Digital Archives

- Das dLZA hat Kriterien für die Auswahl seiner digitalen Objekte entwickelt
  - Das dLZA übernimmt die Verantwortung für den dauerhaften Erhalt der durch die digitalen Objekte repräsentierten Informationen
  - Das dLZA hat seine Zielgruppe(n) definiert
- Das dLZA ermöglicht seinen Zielgruppe(n) eine angemessene Nutzung der durch die digitalen Objekte repräsentierten Informationen
  - Das dLZA ermöglicht seinen Zielgruppe(n) den Zugang zu den durch die digitalen Objekte repräsentierten Informationen
  - Das dLZA stellt die Interpretierbarkeit der digitalen Objekte durch seine Zielgruppe(n) sicher
- Gesetzliche und vertragliche Regelungen werden eingehalten
  - Es bestehen rechtliche Regelungen zwischen Produzenten und dem digitalen Langzeitarchiv
  - Das dLZA handelt bei der Archivierung und Nutzung auf der Basis rechtlicher Regelungen
- Die Organisationsform ist für das dLZA angemessen
  - Die Finanzierung des digitalen Langzeitarchivs ist sichergestellt
  - Es steht Personal mit angemessener Qualifikation in ausreichendem Umfang zur Verfügung
  - Das dLZA betreibt eine langfristige Planung



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture**
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System





# FEDORA – Flexible Extensible Digital Object Repository Architecture



- Flexible Extensible Digital Object Repository Architecture
- A system that provides a digital content repository for a wide variety of users
  - E.g. institutional repository, digital archive, content management system, scholarly publishing enterprises, and digital library
- Fedora as OAIS
  - Is based on OAIS data model, function model and architecture models
  - End client (e.g. CIRILO) can access repository functions via web services



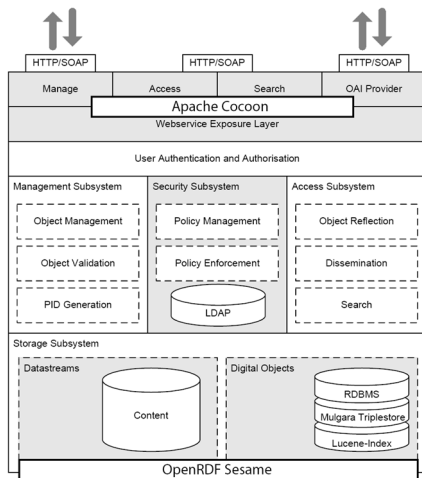
# FEDORA – Flexible Extensible Digital Object Repository Architecture

## Ein Repository zur metadatenbasierten Verwaltung von digitalen Ressourcen

- Entwickelt an der Virginia und der Cornell Universität mit Unterstützung der Mellon Foundation
- Seit 2002 unter Open Source Lizenz verfügbar
- Unterstützt den Aufbau objektorientierter Datenstrukturen
- Webservice-basierte, plattformunabhängige, verteilte Systemarchitektur
- Versionsmanagement mit Integritätsprüfung der geladenen Datenströme
- Volltextindex (Apache Lucene) und RDF-Triplestore (Mulgara)
- Definition von fein granulierbaren Zugriffsrechten auf Basis von XACML
- Unterstützung standardisierter Protokolle zum Datenaustausch: OAI-PMH u.a.
- LDAP- und Shibboleth-basierte Authentifizierung und Autorisierung
- XML basierte Import- und Exportstandards: METS, DIDL u.a.



# FEDORA – Flexible Extensible Digital Object Repository Architecture



# FEDORA – Flexible Extensible Digital Object Repository Architecture

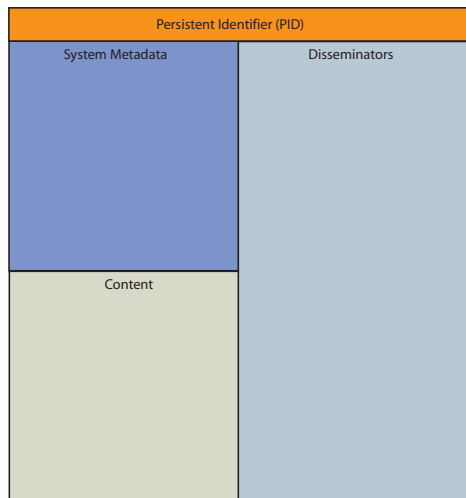
## Datenobjekte und Inhaltsmodelle

- Inhaltsmodelle (Klassen) als Strukturbeschreibung einer zusammengehörigen Menge von digitalen Ressourcen (Dateien, Ergebnisse einer Datenbank- oder Webserviceabfrage u.Ä.) und darauf möglicher Operationen (Disseminatoren, Methoden)
- Kurz: Klassen beschreiben Datenstrukturen und darauf basierende Workflows
- Konkrete Datenobjekte (=Assets) – als Instanziierungen eines Modells (einer Klasse) – weisen damit immer Strukturmerkmale jener Klasse auf, von der sie abgeleitet wurden



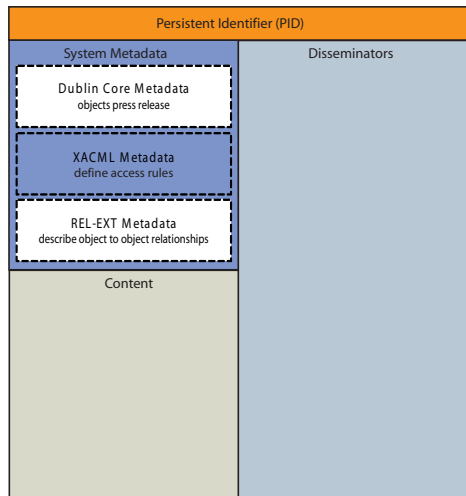
# FEDORA – Flexible Extensible Digital Object Repository Architecture

## A Structural Model of a Content Model



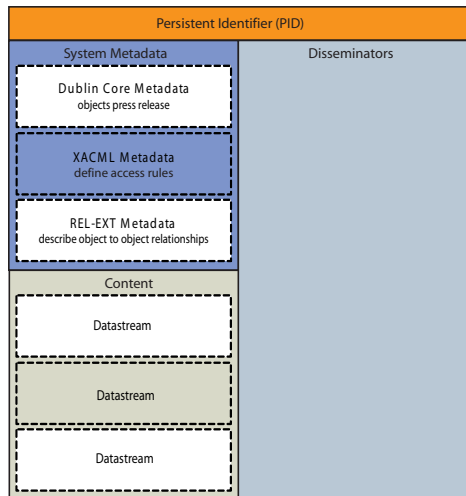
# FEDORA – Flexible Extensible Digital Object Repository Architecture

## A Structural Model of a Content Model



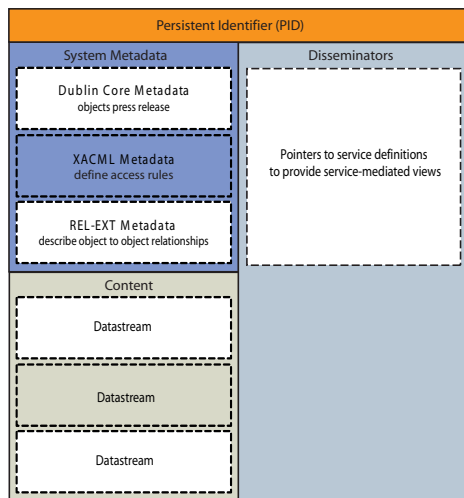
# FEDORA – Flexible Extensible Digital Object Repository Architecture

## A Structural Model of a Content Model



# FEDORA – Flexible Extensible Digital Object Repository Architecture

## A Structural Model of a Content Model





# FEDORA – Flexible Extensible Digital Object Repository Architecture

## Information objects ...

- Shows the datastreams list for the object
  - <http://gams.uni-graz.at/archive/objects/o:mws-115-879/datastreams>
- Shows the methods list for the object
  - <http://gams.uni-graz.at/archive/objects/o:mws-115-879/methods>



# FEDORA – Flexible Extensible Digital Object Repository Architecture

## Projekte basierend auf FEDORA

- eSciDoc – The Open Source e-Research Environment (<https://www.escidoc.org/>)
- TextGrid – Virtuelle Forschungsumgebung für die Geisteswissenschaften (<http://textgridrep.de>)
- PHAIDRA – Permanent Hosting, Archiving and Indexing of Digital Resources and Assets (<https://phaidra.univie.ac.at>)



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen**
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



[illegible][illegible]



DURCH SCHÜLERBÜCHER

Der Aufsatzwelt

Home | Einführung | VDA Auswahlkriterien | Register | DDC & Thesauri

Suche

Logg  
Aus  
Registrieren  
Gast

### Klassische Literaturwissenschaft



Transzendenz über die Mode

1. Auflage, 1978 | PDF | Suchen



PROLOG & EPILOG VON GOETHE

1. Auflage, 1978 | PDF | Suchen

### Aufbauwissenschaften und PoE in DDC



1. Auflage, 1978 | PDF | Suchen



Aufbauwissenschaften und soziale Wirklichkeit

1. Auflage, 1978 | PDF | Suchen

- Chronologische Ordnung
- Aufbau
- Fremde und Gefährungen
- Originalsprache
- Themenkomplex
- Epochen
- Manuskripte

# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

## Geisteswissenschaftliches Asset Management System

- A FEDORA based repository with other embedded open source projects: OpenRDF Sesame and IIP Image Server
- A multitude of web services for content dissemination (e.g. Apache Cocoon)
- Handle System based resolution of persistent identifiers
- A Client named CIRILO for mass ingest and data curation supporting a set of content models
- Workflows and Tools for generating data
- CIRILO and an „archive-in-a-box“ solution as an Austrian contribution to DARIAH  
<https://github.com/acdh/cirilo>



# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System



## Cirilo – Ein Client für FEDORA-basierte Repositories


- Inhaltmodelle kapseln benutzerInnen-spezifische Anforderungen und ermöglichen die einfache Konfiguration projektspezifischer Arbeitsumgebungen
- Unterstützt den Masseneingest von TEI- oder METS-Dokumenten aus Dateisystemen oder eXist-Datenbanken
- Erlaubt das einfache Handling ganzer Objektkollektionen: (z.B. Ersetzungsoperationen über ganze Sammlungen im digitalen Archiv)
- <http://gams.uni-graz.at/cirilo/2.0>

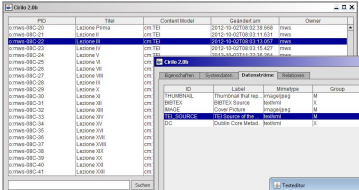


# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

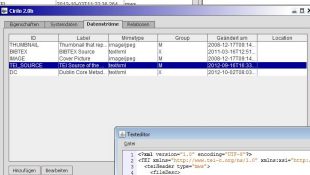
Geisteswissenschaftliches Asset Management System

## Cirilo – Ein Client für FEDORA-basierte Repositories

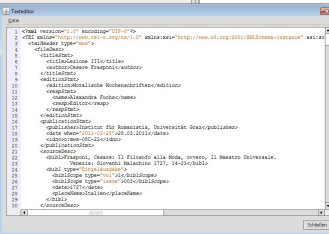
 Cirilo 2.0b  
 Datei Editor Info



PID	Titel	Content Model	Geändert am	Owner
cmwv-abc-20	Lektion I	cm:tei	2012-10-02T08:32:38.668	rwms
cmwv-abc-21	Lektion II	cm:tei	2012-10-02T08:33:11.639	rwms
cmwv-abc-22	Lektion III	cm:tei	2012-10-02T08:33:13.657	rwms
cmwv-abc-23	Lektion IV	cm:tei	2012-10-02T08:33:15.627	rwms
cmwv-abc-24	Lektion V	cm:tei	2012-10-02T08:33:16.664	rwms
cmwv-abc-25	Lektion VI	cm:tei		
cmwv-abc-26	Lektion VII	cm:tei		
cmwv-abc-27	Lektion VIII	cm:tei		
cmwv-abc-28	Lektion IX	cm:tei		
cmwv-abc-29	Lektion X	cm:tei		
cmwv-abc-30	Lektion XI	cm:tei		
cmwv-abc-31	Lektion XII	cm:tei		
cmwv-abc-32	Lektion XIII	cm:tei		
cmwv-abc-33	Lektion XIV	cm:tei		
cmwv-abc-34	Lektion XV	cm:tei		
cmwv-abc-35	Lektion XVI	cm:tei		
cmwv-abc-36	Lektion XVII	cm:tei		
cmwv-abc-37	Lektion XVIII	cm:tei		
cmwv-abc-38	Lektion XIX	cm:tei		
cmwv-abc-39	Lektion XX	cm:tei		
cmwv-abc-40	Lektion XXI	cm:tei		
cmwv-abc-41	Lektion XXII	cm:tei		

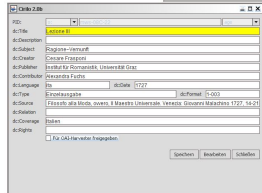


ID	Label	Mimetype	Group	Geändert am	Locales
THUMBNAIL	Thumbnail Thumbnail	image/jpeg	M	2008-12-17T09:14:	
IMTEXT	IMTEXT Source	text/xml	M	2011-03-18T12:51:	
IMAGE	Cover Picture	image/jpeg	M	2009-10-17T08:14:	
TEI SOURCE	TEI Source of the	text/xml	M	2012-06-16T18:35:	
DC	Check Core Values	text/xml	M	2012-10-02T08:33:	



```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/release/5.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:
3   <teiHeader type="text">
4     <fileDesc>
5       <titleDesc>
6         <title>Lektion III</title>
7       </titleDesc>
8       <authorDesc>Georg Franzoni</author>
9     </fileDesc>
10    <editionDesc>
11      <edimstr>Manuskriptschreibung</edimstr>
12    </editionDesc>
13    <groupDesc>
14      <group>Georg Franzoni</group>
15    </groupDesc>
16    <publicationDesc>
17      <publinfo>Institut für Romanistik, Universität Graz</publinfo>
18      <date>2012-10-02T08:33:15.627</date>
19      <class>cmwv-abc-22</class>
20    </publicationDesc>
21    <sourceDesc>
22      <dbl>Franzoni, Georg: II Filodelfo alla Nuda, ovvero, II Maestro Universale.
23        Venezia: Giovanni Marchese 1717, 12-130</dbl>
24      <dbl type="text/plain">
25        <chilscope type="text/plain">
26          <chilscope type="text/plain">
27            <chilscope type="text/plain">
28              <chilscope type="text/plain">
29                <chilscope type="text/plain">
30              </chilscope>
31            </chilscope>
32          </chilscope>
33        </chilscope>
34      </dbl>
35    </sourceDesc>
  
```



PID:	cmwv-abc-22
dc:title:	Lektion III
dc:description:	Manuskriptschreibung
dc:subject:	Georg Franzoni
dc:creator:	Institut für Romanistik, Universität Graz
dc:publisher:	Georg Franzoni
dc:contributor:	Georg Franzoni
dc:language:	ita
dc:type:	Manuskriptschreibung
dc:format:	text/xml
dc:source:	Filodelfo alla Nuda, ovvero, Il Maestro Universale. Venezia: Giovanni Marchese 1717, 12-130
dc:relation:	Teil
dc:coverage:	italien
dc:rights:	<input type="checkbox"/> Public Domain





# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

- Digitales Archiv (Repository) zur Verwaltung und Bereitstellung von heterogenen, digitalen Ressourcen auf Basis von FEDORA (Flexible Extensible Digital Object Repository Architecture)
- Das FEDORA-Framework unterstützt die flexible und nachhaltige Nutzung digital vorliegender Inhalte durch ein objektorientiertes Konzept sogenannter „Inhaltsmodelle“
- Research Driven Development: System muss offen und flexibel auf (u.U. noch nicht bekannte) Bedarfe reagieren können.



# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

- Konsequente Umsetzung des XML-Credos: Aus SIPs mach AIPs
- Jedes Objekt im Repository verfügt über das nötige „Wissen“, um sich selbst zu kontextualisieren und zu präsentieren
- Objekt-Präsentationen werden in der Regel on-the-fly, als Antwort auf einen Request aus dem Inhalt des Objektes generiert
- System-Metadaten werden automatisiert nach (projektspezifischen) Regelwerken aus den Inhalten extrahiert



# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

## Inhaltsmodelle

Persistent Identifier (PID)		
System-Metadaten	Inhalte	Disseminatoren

# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

## Inhaltsmodelle

Persistent Identifier (PID)		
<b>System-Metadaten</b> <div> <div>Deskriptive Metadaten</div> <div>Semantische Relationen des Objektes</div> <div>Zugriffsrechte auf das Objekt</div> <div>Änderungshistorie des Objektes</div> </div>	Inhalte	Disseminatoren

# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

## Inhaltsmodelle

Persistent Identifier (PID)		
System-Metadaten	Inhalte	Disseminatoren
Deskriptive Metadaten	XML-Datenströme	
Semantische Relationen des Objektes	Daten, Bilder, Videos, Audios u.a.	
Zugriffsrechte auf das Objekt	XSLT-Stylesheets zum Erstellen von (Re-)Präsentationen des Objekthinhaltes	
Änderungshistorie des Objektes		

# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

## Inhaltsmodelle

Persistent Identifier (PID)		
System-Metadaten	Inhalte	Disseminatoren
Deskriptive Metadaten	XML-Datenströme	Methoden zum Zugriff auf Objekthinhalte
Semantische Relationen des Objektes	Daten, Bilder, Videos, Audios u.a.	
Zugriffsrechte auf das Objekt	XSLT-Stylesheets zum Erstellen von (Re-)Präsentationen des Objekthinhaltes	
Änderungshistorie des Objektes		

# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

## Skizze des Strukturmodells eines TEI-Objektes – Datenströme

- Inhalt
  - THUMBNAIL
  - DC
  - TEI\_SOURCE
  - BIBTEX
  - ...
- System
  - STYLESHEET
  - FO\_STYLESHEET
  - DC\_MAPPING
  - RDF\_MAPPING
  - ...

## Disseminatoren

- sdf:Object
  - get
  - getDC
  - getMetadata
- sdf:TEI
  - get
  - getPDF
- sdf:BibTeX
  - get
  - getBIBTEX
  - getRIS
  - getENDNOTE
  - getMODS
- sdf:Vovant
  - get



# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

## Das alles löst der Upload eines TEI-Dokumentes im Repository aus

- Regelbasierte Extraktion von deskriptiven Metadaten aus den Primärdaten
- Regelbasierte Extraktion von semantischen Informationen aus den Primärdaten
- Auflösen der Kodierung von Konzepten aus kontrollierten Vokabularen (z.B. SKOS-Objekt)
- Anwendung eines Annotationsregelwerks, z.B. zur automatisierten Auszeichnung von Namen, Ortsbezeichnungen u.Ä. auf Basis von regulären Ausdrücken
- Kodierungsgesteuerte Zuordnung des Objekts zu Sammlungen und anderen Kontexten im Archiv
- Kodierungsgesteuerter Upload von Bilddateien
- Validierung der Bilddateien und Berechnung von Bildvarianten (Thumbnails, niederauflösende Varianten u.Ä.)
- Automatische Extraktion von technischen Metadaten aus Bilddateien





# GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen

Geisteswissenschaftliches Asset Management System

## Beispiel: Inhaltsmodell der Montfort-Edition

Ein Folio-Objekt der Hugo von Montfort-Edition kapselt sämtliche für die Repräsentation einer Handschriftenseite im Browser notwendigen Datenströme und Funktionalitäten (Bilder der Foliospalten, transliterierter Text der Handschrift, aber auch Zugriffsmethoden, wie z.B. die Lupe).

Persistent Identifier (PID)		
Metadaten	Inhalt	Disseminatoren
DCM Default Schema unter FEDORA für OAI Harvester u.Ä.	TEL-kodierte Transliteration einer Folioseite inkl. textkritischer Anmerkungen	On-The-Fly Transformation der Objekthinhalte zur Anzeige in einem Browser
REL-EXT beschreibt Relationen zu anderen Objekten im Repository	Faksimilies (der einzelnen Spalten) einer Folioseite	On-The-Fly Transformation der Objekthinhalte zu einem PDF-Dokument
XACML definiert Zugriffsrechte auf Inhalte und Disseminatoren des Objektes	Referenzen auf XSLT-Stylesheets u.Ä.	...



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model**
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



# OAIS – Open Archival Information System Reference Model

- In 2000 the Research Libraries Group (RLG) and Online Computer Library Center (OCLC) discussed how both organizations could build an infrastructure for purposes of archiving digital objects.
- The resulting model guides you through building an archival information system



# OAIS – Open Archival Information System Reference Model

## Prämissen der Philosophie

- Dient zur Standardisierung von Bearbeitungs-, Ingest- und Migrationsabläufe - allgemein Workflows - innerhalb von digitalen Archiven
- Wurde 2002 als ISO-Norm eingeführt
- Amerikanische nationale Archivverwaltung (NARA) und die Research Libraries Group (RLG) waren maßgeblich an der Definition des Standards beteiligt
- Zielt in erster Linie auf die Standardisierung von Abläufen im Bereich der Langzeitarchivierung digitaler Bestände
- Fokussiert dabei auf die Langzeitaufbewahrung und Langzeitnutzbarmachung (hauptsächlich) digital vorliegender Bestände und dies unter der Berücksichtigung sich verändernder Technologien
- OAIS verzichtet auf Beschränkungen hinsichtlich bestimmter Systemarchitekturen
- Versucht die klassischen archivischen Arbeitsfelder Erfassen, Aussondern, Bewerten, Übernehmen, Erschließen, Erhalten und Zugänglichmachen abzudecken
- Im Prinzip beantwortet das OAIS-Referenzmodells die schon häufig gestellte Frage nach dem zukünftigen Aufgabenspektrum von Museen, Archiven, Bibliotheken und anderen einschlägigen Institutionen im digitalen Zeitalter



# OAIS – Open Archival Information System Reference Model

## Prämissen der Philosophie

- Referenzmodell für ein dynamisches, erweiterbares Archivsystem mit Anspruch auf Allgemeingültigkeit
- Verzichtet auf eine Beschränkung bezüglich Datentypen, Datentypen oder Systemarchitekturen (im technischen Sinne)
- Beschreibt funktional die Tätigkeitsfelder eines OAIS-Betreibers und damit Kernbereiche notwendiger Operationen, die für die dauerhafte Aufbewahrung digitaler Aufzeichnungen und für die Gewährleistung des Zuganges zu ihnen unverzichtbar sind
- OAIS beschreibt ein Informationsnetzwerk, dass den/die ArchivarIn und den/die NutzerIn als Hauptkomponenten des digitalen Archivs versteht
- Grundsatz: Archivierung ist nicht ausschließlich etwas, was an Maschinen delegierbar ist
- OAIS berücksichtigt immer auch die Verknüpfung von analogen und digitalen Repräsentationen von Entitäten im Archiv



# OAIS – Open Archival Information System Reference Model

## Defines a long-term preservation terminology for

- Architectures and Operations
- Preservation strategies and techniques
- A data model

## Provides ...

- a framework for an increased awareness of concepts needed for long term preservation
- a framework for describing and comparing architectures and operations of archives
- a basis for comparing data models of digital information preserved by archives



# OAIS – Open Archival Information System Reference Model

## Roles

- Producer – A data provider
- Administrator – A data manager
- Consumer – A data retriever

## Important functions

- Ingest – Submit data to system
- Store – Preserve data in system
- Access – Retrieve data from system



# OAIS – Open Archival Information System Reference Model

## Roles and their responsibilities

- **Producer**
  - Person(s) or client system(s), who provide the information to be preserved
  - Ingest digital resource to system
- **Administrator**
  - Person(s) or client system(s), who manage and set the overall policy of the OAIS
  - Monitor, verify digital resource, do preservation planning, migrate digital resource, and etc.
- **Consumer**
  - Person(s) or client system(s), who interact with the OAIS system and services
  - Search and access digital resource in repository





# OAIS – Open Archival Information System Reference Model

## Data Model

- Preserved data in the system needs to be wrapped in a package
- Owing to the three important functions of OAIS (Ingest, Store, and Access), packages of preserved data are transformed into three types
  - Submission Information Package
  - Archival Information Package
  - Dissemination Information Package



# OAIS – Open Archival Information System Reference Model

## Kernkomponenten: Informationsobjekte und Datenmodell

OAIS unterscheidet zwischen drei sogenannten Informationsobjekten, die miteinander in Verbindung stehen und sich aufeinander beziehen

- Submission Information Packages (SIP), sind jene Informationseinheiten, die das Archiv an digitalen Unterlagen zur Verwahrung übernimmt
- Archival Information Packages (AIP), SIPs werden in Archiv zu AIPs umgeformt, indem Metainformationen ergänzt werden und die Inhalte in Formen gebracht werden, so dass sie tatsächlich langfristig aufbewahrt werden können
- Dissemination Information Packages (DIP), machen AIPs für unterschiedliche BenutzerInnengruppen zielgruppenorientiert (z.B. unter Berücksichtigung verschiedener rechtlicher Rahmenbedingungen) zugänglich. Bruch mit dem klassischen Verständnis: BenutzerIn wird vorgelegt, was im Magazin verwahrt wird.



# OAIS – Open Archival Information System Reference Model

## Kernkomponenten: Informationsobjekte

Einzelne Informationspakete werden im Rahmen von OAIS als digitale Objekte verstanden. Jedes Informationspaket enthält

- Inhaltliche Informationen (Content Information), die aus den übernommenen und ggf. aufbereiteten Ursprungsdaten bestehen und den beschriebenen Repräsentationsformen bestehen
- Informationen zur Beschreibung der Aufbewahrungsform (Preservation Description Information), sie beschreiben, was an Technik und Verfahren angewandt wurde und sie zu sichern und langfristig nutzbar zu machen



# OAIS – Open Archival Information System Reference Model

## Preservation Description Information – PDI

**Provinienz** Beschreibung der Quelle der Inhaltsinformationen eines Objektes und dessen Objekthistorie

**Kontext** Informationen die Verbindungen von Inhaltsinformationen mit anderen Informationen außerhalb des Objektes beschreiben

**Beziehungen** System von Signaturen machen Informationspakete eindeutig identifizierbar und unterscheidbar

**Stabilisierung** Inhaltsinformationen die die Authentizität des Objektes sicherstellen. Objekt muß vor unbeabsichtigter Veränderung bewahrt werden



# OAIS – Open Archival Information System Reference Model

## PDI – Example of PDI Information

- A Digital Library Collection
  - Provenance
    - For scanned collections: metadata about the digitisation process, pointer to master version
    - For born digital publications: Pointer to digital original
    - Metadata about the preservation process: pointer the earlier versions, change history
  - Context
    - Pointer to related documents in original environment at the time of publication
  - Reference
    - Bibliographic Description
    - Persistent Identifier
  - Fixity
    - Digital Signature
    - Checksum
    - Authenticity Indicator



# OAIS – Open Archival Information System Reference Model

## PDI – Example of PDI Information

- Software Package
  - Provenance
    - Revision History
    - License Holder
    - Registration
    - Copyright
  - Context
    - Help file
    - Users Guide
    - Related software
  - Reference
    - Name
    - Author
    - Version Number
    - Serial Number
  - Fixity
    - Certificate
    - Checksum



# OAIS – Open Archival Information System Reference Model

## Administrator

- Negotiate Submission Agreement
  - Discuss submission agreement with producer
- Manage System Configuration
  - Configure and control changes which affect system engineering of archival system
- Physical Access Control
  - Authorize access to resources
- Establish Standards and Policies
  - Manage standards and policies in order to approve migration and replication processes
- Audit Submission
  - Verify that AIP and SIP is following specification and agreement
- Activate Requests
  - To check the request of consumer is correct, then submit the request to Access
- Customer Service
  - Provide functions to manage user account



# OAIS – Open Archival Information System Reference Model

## Administrator (cont.) – Preservation Planning

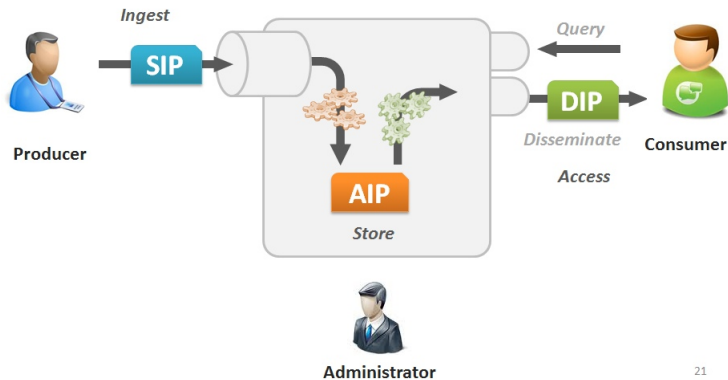
- Monitor environment of OAIS and provide recommendations
- Monitor Designated Community
  - Allow consumer and producer to track change of available technologies
- Monitor Technology
  - Report change of software and hardware contributing to preservation process
- Develop Preservation Strategies and Standards
  - Develop and recommend strategies and standards for future change of technology
- Develop Packaging Designs and Migration Plans
  - Customize SIP and AIP template for migration goal





# OAIS – Open Archival Information System Reference Model

3 functions, 3 information packages, and 3 roles



21

# OAIS – Open Archival Information System Reference Model

## Ingest

- Accept SIPs from Producers
  - Verify SIPs that user submits
  - Generate AIPs for archive storage
- 
- Receive Submission
    - Upload SIP package from producer by electronic transfer such as FTP
  - Quality Assurance
    - Validate transmission (e.g. checksum) error SIP package and log a result
  - Generate AIP
    - Transform SIP to AIP and report result
  - Coordinate Update
    - Provide a single access point (add, modify, remove, get) to storage area



# OAIS – Open Archival Information System Reference Model

## Store

- Receive Data
  - Receive AIP from Ingest to permanent storage
- Manage Storage Hierarchy
  - Provide administration functions for storage media
- Replace Media
  - Support functions of migration from a media to another media
- Error Checking
  - Check and notification error from data in storage area
- Disaster Recovery
  - Provide mechanism for replicating digital content to backup



# OAIS – Open Archival Information System Reference Model

## Access

- Coordinate Access Activities
  - Provide single user interface for features like browse, search and access
- Generate DIP
  - Generate DIP from AIP
- Generate DIP
  - Handle response from query and access and delivery to consumer
  - Report access activities to administrator



# OAIS – Open Archival Information System Reference Model

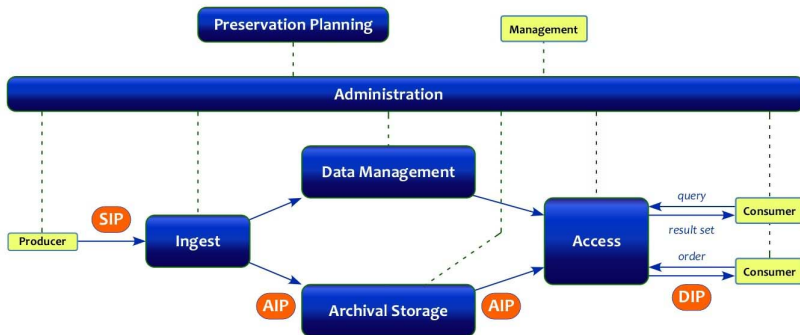
## Summary

- Producer
  - Ingest package to system
    - System stores AIP in archival storage
    - System stores descriptive metadata in data management
- Consumer
  - Queries data via Access
    - Query from descriptive metadata from data management
  - Retrieve data via access
    - Get data from archival storage
- Administrator
  - Manages and monitors every process in system



# OAIS – Open Archival Information System Reference Model

Big picture of all functions and process of packages



**OAIS Reference Model**



# OAIS – Open Archival Information System Reference Model

## Useful Resources

- OAIS specification
  - <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- OAIS primer
  - [http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf)



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies**
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System





# PREMIS – Preservation Metadata: Implementation Strategies

## Grundlegendes Prinzip: Technikneutralität

- Definiert ein allgemeines Datenmodell zur Organisation von Preservation Metadata
- Beschreibt den Life Cycle eines digitalen Objektes
- Keine Annahme über spezifische digitale Archivsysteme oder -technologien
- Keine Annahme über bestimmte Formen des Metadatenmanagements
- Ist Referenzmodell für den Austausch von Informationspaketen zwischen unterschiedlichen Archivsystemen
- Keine fertige „Out-of-the-box“ Lösung



# PREMIS – Preservation Metadata: Implementation Strategies

## Ziele

- Metadatenschema für Informationen zur Langzeitarchivierung
- Leitfaden für die Entwicklung oder Bewertung von Softwareanwendungen
- Informationen, die einem Langzeitarchiv „bekannt“ sein sollten, um seine Archivierungsfunktion erfüllen zu können
- Informationen über digitale Objekte, ihre Beziehungen und Verknüpfungen, welche Prozesse der digitalen Langzeitarchivierung ermöglichen, unterstützen und dokumentieren
- Exemplarische Aufgabenbereiche in Langzeitarchivierungskontexten
  - Sicherstellung der Authentizität von Daten (bspw. durch Speicherung von Checksummen oder Dokumentation der Provenienz, der Verarbeitungskette einer Ressource)
  - Sicherstellung eines automatischen Refreshings von Datenträgern
  - Sicherstellung der automatisierten Migration von Dateiformaten



# PREMIS – Preservation Metadata: Implementation Strategies

## Different Types of Metadata

- Descriptive
  - Supports the identification and discovery of a resource
- Administrative
  - Supports the management and tracking of a resource
- Structural
  - Defines the arrangement and composition of a resource
- Preservation
  - Supports activities intended to ensure the long term usability of a resource



# PREMIS – Preservation Metadata: Implementation Strategies

## Expizite Nicht-Ziele

- Formatspezifische Metadaten
- Metadaten, die anwendungs- oder geschäftsablaufspezifisch sind, z.B. die Policy oder die Praxis eines individuellen Archivs betreffend, z.B. in Bezug auf die Zugänglichmachung von digitalen Inhalten
- Deskriptive Metadaten, hierzu gibt es genügend etablierte Standards
- Detaillierte Informationen über Speichermedien oder Hardware
- Informationen über Rechte und Genehmigungen, bis auf jene die sich unmittelbar auf die Langzeitarchivierung auswirken



# PREMIS – Preservation Metadata: Implementation Strategies

## Semantische Einheiten

- PREMIS definiert keine Metadatenelemente, sondern semantische Einheiten
- Eine semantische Einheit ist ein Stück Information oder Wissen
- Ein Metadatenelement ist eine definierte Vorgabe für die Darstellung von Informationen
- PREMIS definiert nicht wie Metadaten in einem System repräsentiert werden sollen, sondern bloß, was ein System wissen soll und was es in andere Systeme exportieren können soll
- Semantische Einheiten können natürlich direkt in Metadatenelemente, die im PREMIS XML-Schema definiert sind, aber auch andere Metdatenschema gemappt werden
- PREMIS ist ein offizielles Erweiterungsschema von METS



# PREMIS – Preservation Metadata: Implementation Strategies

## Das Datenmodell

- Intellektuelle Entitäten** Sie sind unter PREMIS konzeptionell und werden als ein Set von Inhalten betrachtet, die zu Zwecken des Managements und der Beschreibung als eine Einheit zu sehen sind (bspw. ein Buch, eine Landkarte, eine Datenbank). Eine Intellektuelle Einheit kann andere IEs enthalten (bspw. eine Website eine Webseite)
- Objekte** Das was tatsächlich in einem digitalen Objekt gespeichert und verwaltet wird
- Ereignisse** Diese Ebene des Datenmodells führt Informationen über Handlungen zusammen, die die Objekte im digitalen Langzeitarchiv betreffen
- Rechte** Diese Ebene fasst Informationen über rechtliche Angaben zusammen, die unmittelbar für die Erhaltung von Objekten in einem Langzeitarchiv von Bedeutung sind (z.B. bezüglich des Anfertigungs von identen Kopien u.Ä.)
- Agenten** Agenten sind Handelnde (Personen, Organisationen, Algorithmen) , die in Bezug auf Ereignisse oder Rechte eine Rolle spielen. Hier gibt es unter PREMIS nur eine Minimalzahl an semantischen Einheiten, da es bereits eine Vielzahl von eingeführten Standards gibt



# PREMIS – Preservation Metadata: Implementation Strategies

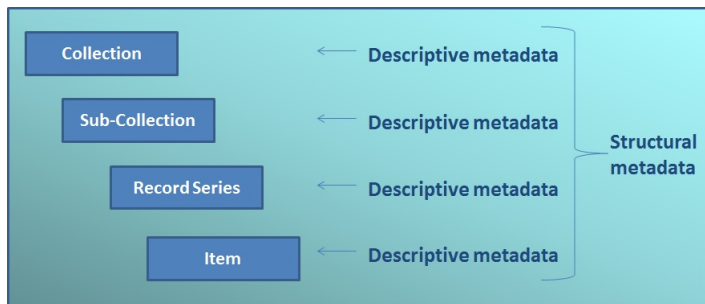
## Usage

- Repository Design
  - Provides guidelines on what information should be obtained and maintained by a preservation repository
- Repository evaluation
  - Provides a checklist to determine effective preservation management of digital objects
- Exchange of objects between repositories
  - Provides a common set of data elements that can be understood by the provider and consumer repositories



# PREMIS – Preservation Metadata: Implementation Strategies

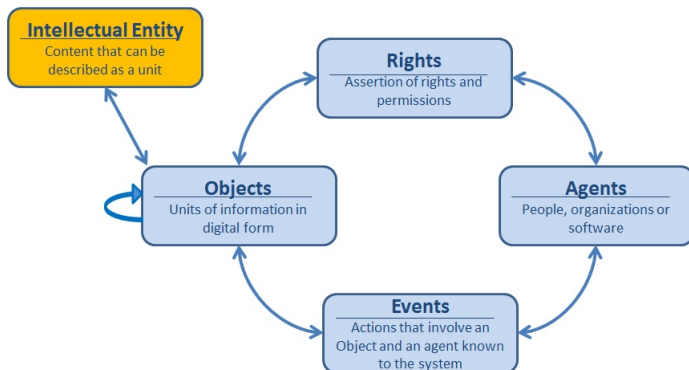
Always has intellectual entities





# PREMIS – Preservation Metadata: Implementation Strategies

## Data Model



# PREMIS – Preservation Metadata: Implementation Strategies

## Semantic Units

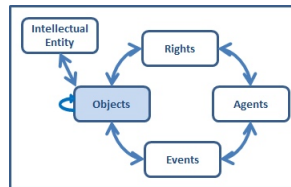
- Semantic Units
  - Convey a piece of information / knowledge
  - Do not specify how they should be represented in a particular system (e.g. to metadata elements)
  - Should be exportable to other systems
  - May have a direct mapping to metadata elements in an XML schema
- Repository evaluation
  - Provides a checklist to determine effective preservation management of digital objects
- Containers and sub units
  - Some semantic units are defined as container
  - Facilitates a hierarchical structure to the data dictionary



# PREMIS – Preservation Metadata: Implementation Strategies

## Objects

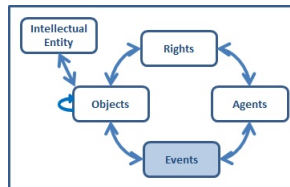
- Identifier
- Category (Representation, File, Bitstream)
- Preservation level
- Characteristics: Fixity, Size (bytes), Format, Creating application
- Original name
- Environment: ...,Software, Hardware, ...
- Signature Information
- Linked events
- Linked intellectual entity
- Linked rights statement



# PREMIS – Preservation Metadata: Implementation Strategies

## Events

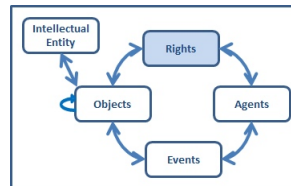
- Identifier
- Type
- Date Time
- Outcome Information
- Linking Agent Identifier
- Linking Object Identifier



# PREMIS – Preservation Metadata: Implementation Strategies

## Rights

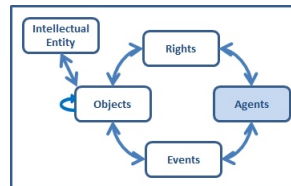
- Identifier
- Copyright Information
- Licence Information
- Rights granted
- Linking Agent Identifier
- Linking Object Identifier



# PREMIS – Preservation Metadata: Implementation Strategies

## Agents

- Identifier
- Name
- Type



# PREMIS – Preservation Metadata: Implementation Strategies

## Useful Resources

- PREMIS specification
  - <http://www.loc.gov/standards/premis/>
- PREMIS primer
  - <http://www.loc.gov/standards/premis/understanding-premis.pdf>



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative**
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System





# DCMI – Dublin Core Metadata Initiative

## Ziele

- Einfachheit
- Erreichung größerer Recherchepräzision als bei der gewohnten Volltextsuche
- Orientiert sich an der Prämisse, dass der/die AutorIn LieferantIn der Metadaten sein können muß
- Allgemein akzeptiertes, einfaches Metadaten-set zur Beschreibung von Internetressourcen
- Soll mit Metadaten-sets in anderen Formaten verknüpfbar und damit erweiterbar sein
- Dublin Core Metadata Element Set ist in verschiedenen Syntaxformaten darstellbar (XML, RDF u.a.)



# DCMI – Dublin Core Metadata Initiative

## Elemente und Qualifier

- Dublin Core Namensraum mit Präfix dc: dargestellt, URL <http://purl.org/dc/elements/1.0>
- Der Kernsatz des DC besteht aus 15 Elementen
- Optionale Qualifier verfeinern (Refinement) oder interpretieren den Wert eines Elementes mit Hilfe von Vokabularen
- Die Verfeinerungen (Dublin Core Metadata Terms) werden mit dem Namensraumpräfix dcterms: dargestellt, URL <http://purl.org/dc/terms>
- Alle Elemente haben eine Kardinalität von 0 .. n
- Zusätzliche Standardisierung bringt die Verwendung von Vocabulary Encoding Schema (z.B. DCMI Type Vocabulary, Getty Thesaurus of Geographic Names uvm.)



# DCMI – Dublin Core Metadata Initiative

## Elemente des Core Sets

- dc:title** Titel einer Quelle; der vom Verfasser, Urheber vergebene Name der Ressource
- dc:creator** Verfasser oder Urheber; Personen oder Organisationen, die den intellektuellen Inhalt verantworten
- dc:subject** Schlagwort, Stichwort das die Ressource inhaltlich erschließt
- dc:description** Textuelle Beschreibung des Ressourceninhaltes
- dc:publisher** Verleger, Herausgeber; Person oder Einrichtung, die verantwortet, daß die Ressource zur Verfügung steht
- dc:contributors** Zusätzliche Personen oder Körperschaften die sekundär intellektuelle Beiträge sind
- dc:date** Datum an dem die Ressource in der gegenwärtigen Form zugänglich gemacht wurde



# DCMI – Dublin Core Metadata Initiative

## Elemente des Core Sets

- dc:type** Ressourcenart, z.B. Homepage, Roman, Gedicht, Arbeitsbericht, Wörterbuch ...
- dc:format** Bezeichnet das datentechnische Format der Ressource, z.B. Text/HTML, JPEG
- dc:identifier** Eindeutiger Kenner der Ressource
- dc:source** Quelle des Inhaltes der digitalen Resource
- dc:language** Sprache des intellektuellen Inhalts der Ressource
- dc:relation** Angabe des Bezugs zu anderen Ressourcen
- dc:coverage** Angaben zur räumlichen Bestimmung und zeitlichen Gültigkeit
- dc:rights** Verweis auf rechtliche Benutzungsbedingungen



# DCMI – Dublin Core Metadata Initiative

## Elemente aus Metadata Terms

- dc:title = dcterms:alternative
- dc:description = dcterms:abstract, dcterms:tableOfContents
- dc:date = dcterms:available, dcterms:created, dcterms:issued, dcterms:modified
- dc:coverage = dcterms:spatial, dcterms:temporal
- dc:rights = dcterms:accessRights, dcterms:rightsHolder, dcterms:license
- dc:relation = dcterms:conformsTo, dcterms:isPartOf, dcterms:isReferencedBy, dcterms:isReplacedBy, dcterms:isRequiredBy, dcterms:isVersionOf
- dc:format = dcterms:extent



# DCMI – Dublin Core Metadata Initiative

## Beispiel in XML-Notation

```
<oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
  <dc:title>XML-Frameworks im Korpusmanagement</dc:title>
  <dc:creator>Johannes Stigler</dc:creator>
  <dc:identifier>o:gralis-074-1</dc:identifier>
  <dc:date>2008</dc:date>
  <dc:language>deu</dc:language>
  <dc:publisher>Zentrum für Informationsmodellierung, Universität Graz</dc:publisher>
  <dc:subject>Korpuslinguistik</dc:subject>
  <dc:subject>Informationsmodellierung</dc:subject>
</oai_dc:dc>
```



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard**
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



# METS – Metadata Encoding and Transmission Standard

## Allgemeines

- Containerformat für Dokumente in das Inhalte beliebigen Formats eingebunden werden können.
- METS beschreibt auch das Verhältnis der Dokumente zueinander
- Kann logische und physische Strukturen eines zu beschreibenden Dokumentes abbilden und bietet explizite Möglichkeiten diese beiden Strukturebenen miteinander zu verknüpfen
- Neben Strukturen berücksichtigt das Modell auch deskriptive (bibliographische) und administrative (Informationen zu Nutzungsrechten, technische Informationen zu einzelnen Dateien u.Ä.) Metadaten
- Es lassen sich sowohl Inhalte beliebiger Metadatenschema als auch Text und Binärinhalte einbinden
- METS wird derzeit fast ausschließlich als XML serialisiert und in Form von Dateien gespeichert





# METS – Metadata Encoding and Transmission Standard

## Dateien spielen eine wesentliche Rolle

- Entsprechende Daten können direkt (auch BASE64 encoded) oder als Referenz in ein METS-Dokument eingebunden werden (bspw. Transkription im TEI-Format, Images gescannter Buchseiten)
- Ein METS-Dokument ist also ein Container für alle für ein Dokument notwendigen Dateien, unabhängig davon, ob die Dateien lokal oder auf dislozierten Servern vorhanden sind
- Dateien lassen sich gruppieren und z.B. nach Verwendungszwecken (bspw. technische Metadaten) einzelnen Gruppen zuordnen
- Weiters ist es möglich die Versionsgeschichte von Dateien zu modellieren
- Das METS-Abstract Model erlaubt das Speichern von administrativen Metadaten (z.B. Dateigröße, Checksummen u.Ä.) zu jeder Datei
- Dateien können darüber hinaus mit Struktureinheiten (mittels M:N Relation) verknüpft werden (bspw. Imageseite mit Inhaltsverzeichnis)
- Auch zeitliche Abschnitte eines Streams lassen sich mit logischen Strukturen (z.B. Navigationsstruktur auf einen Videostream) eines Dokumentes verknüpfen



# METS – Metadata Encoding and Transmission Standard

## Dateien spielen eine wesentliche Rolle

- Die Flexibilität des Formates erfordert es eine Dokumentation der konkreten Implementierung von METS zu erstellen
- Das METS Profil-Schema bietet eine standardisierte Möglichkeit, eine solche Dokumentation zu erstellen
- Ein solches Profil kann durch das METS-Editorial-Board verifiziert werden und steht dann auf der Homepage zur Nachnutzung bereit
- Neben der ursprünglichen Anwendung, digitalisierte Bücher zu beschreiben, wird es heute sowohl für Webseitenbeschreibungen als auch ganz allgemein zur Beschreibung von komplexen Dokumenten verwendet und bekommt gerade im Kontext der Langzeitarchivierung als De-Facto-Standard immer größere Bedeutung



# METS – Metadata Encoding and Transmission Standard

## Ein METS-Dokument besteht aus 7 Hauptabschnitten

- Kopfteil: METS Header
- Erschließungsangaben: Descriptive Metadata
- Verwaltungsangaben: Administrative Metadata
- Dateiabschnitt: File Section
- Strukturbeschreibung: Structural Map
- Strukturverknüpfungen: Structural Links
- Verhaltenbeschreibung: Behavior



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework**
- 11 SKOS – Simple Knowledge Organization System



# RDF – Resource Description Framework

## RDF

- XML ist geeignet als Datenformat bietet aber keine Standards für die Verarbeitung
- Standardisierung von Semantik fehlt. Syntaxverschiedenheit bei semantisch identen Ausdrücken ist möglich  
Wie modelliert man/frau in XML die Aussage: „Das ZIM ist eine Subeinheit der Universität Graz?“

```
<department>
  <name>ZIM</name>
  <org>Universität Graz</org>
</department>

<department name="ZIM">
  <org>Universität Graz</org>
</department>

<org name="Universität Graz">
  <department>ZIM</department>
</org>
```

- RDF als Vorschlag des W3C zum Austausch von maschineninterpretierbaren Metainformationen
- RDF ist grundsätzlich erweiterbar und erlaubt die Verwendung mehrerer Schemata



# RDF – Resource Description Framework

## Grundlagen

- RDF ist ein Modell zur Beschreibung von Ressourcen
- Ressourcen sind materielle oder immaterielle Objekte, die über einen URI eindeutig identifiziert werden können
- Eigenschaften sind Merkmale welche die Ressource beschreiben
- Eine RDF-Aussage setzt sich aus einer Ressource (Subjekt), einer benannten Eigenschaft (Prädikat) und dem Wert der Eigenschaft (Objekt) zusammen
- RDF-Aussagen werden auch als auch als *Triple* (S,P,O) bezeichnet

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ex="http://example.org">
  <rdf:Description rdf:about="ZIM">
    <ex:isDepartmentOf>Universität Graz</ex:isDepartmentOf>
  </rdf:Description>
</rdf:RDF>
```



# RDF – Resource Description Framework

## Grundlagen

- `http://gams.uni-graz.at/fedora/get/o:gralis-074-1` (Subjekt) hat Autor (Prädikat) „Johannes Stigler“ (Objekt)

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://gams.uni-graz.at/fedora/get/o:gralis-074-1">
    <dc:creator>Johannes Stigler</dc:creator>
  </rdf:Description>
</rdf:RDF>
```



# DCMI – Dublin Core Metadata Initiative

## Beispiel in RDF-Notation

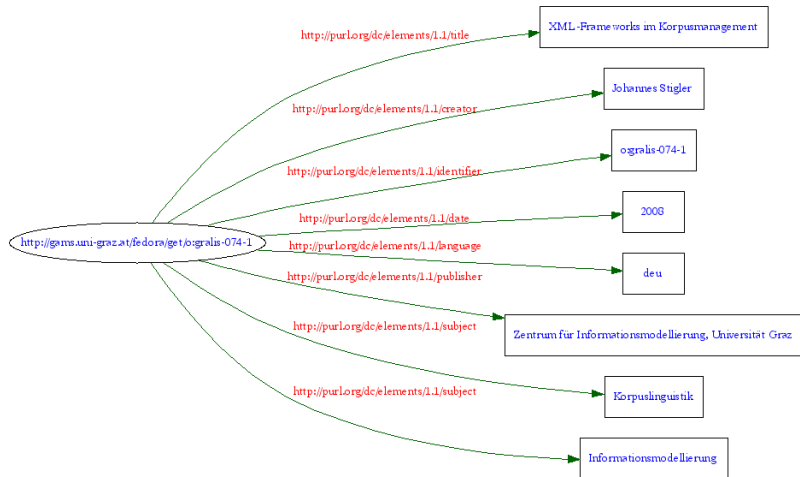
```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://gams.uni-graz.at/fedora/get/o:gralis-074-1">
    <dc:title>XML-Frameworks im Korpusmanagement</dc:title>
    <dc:creator>Johannes Stigler</dc:creator>
    <dc:identifier>o:gralis-074-1</dc:identifier>
    <dc:date>2008</dc:date>
    <dc:language>deu</dc:language>
    <dc:publisher>Zentrum für Informationsmodellierung, Universität Graz</dc:publisher>
    <dc:subject>Korpuslinguistik</dc:subject>
    <dc:subject>Informationsmodellierung</dc:subject>
  </rdf:Description>
</rdf:RDF>
```





# RDF – Resource Description Framework

## Ein Dublin Core Metadata Set als RDF-Graph



# RDF – Resource Description Framework

- Die Art der Beschreibung einer Ressource ist von deren Informationsinhalt abhängig. Kann z.B. also auch gemäß Dublin Core Schema modelliert sein
- Eigenschaften können näher beschreiben werden und ihrerseits wiederum RDF-Aussagen sein
- Die Person unter <https://...?PID=6FAA7A93B3B3AD57> ist „Johannes Stigler“ und hat die eMail-Adresse „johannes.stigler@uni-graz.at“. Er ist der Autor von <http://gams.uni-graz.at/fedora/get/o:gralis-074-1>

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:vCard="http://imc.org/vCard/3.0#">
  <rdf:Description rdf:about="http://gams.uni-graz.at/fedora/get/o:gralis-074-1">
    <dc:creator>
      <rdf:Description rdf:about="https://...?PID=6FAA7A93B3B3AD57">
        <vCard:FN>Johannes Stigler</vCard:FN>
        <vCard:EMAIL>johannes.stigler@uni-graz.at</vCard:EMAIL>
      </rdf:Description>
    </dc:creator>
  </rdf:Description>
</rdf:RDF>
```



# Agenda

- 1 Organisatorisches
- 2 Software
- 3 Allgemeines
- 4 FEDORA – Flexible Extensible Digital Object Repository Architecture
- 5 GAMS – Ein Strukturprojekt zur Verwaltung digitaler Ressourcen
- 6 OAIS – Open Archival Information System Reference Model
- 7 PREMIS – Preservation Metadata: Implementation Strategies
- 8 DCMI – Dublin Core Metadata Initiative
- 9 METS – Metadata Encoding and Transmission Standard
- 10 RDF – Resource Description Framework
- 11 SKOS – Simple Knowledge Organization System



# SKOS – Simple Knowledge Organization System

- SKOS wurde entwickelt um kontrollierte Vokabulare (Taxonomien, Thesauri u.a.) Linked-Data-konform zu modellieren
- Grundbausteine sind Konzepte, denen Notationen, Bezeichnungen, Kommentare uvm. zugeordnet werden können
- Konzepte können anhand hierarchischer und assoziativer Relationen miteinander verknüpft werden
- Neben den Relationen, die innerhalb eines kontrollierten Vokabulars bestehen, können zusätzlich Verbindungen zu anderen Vokabularien etabliert werden
- Best Practice Beispiel: Standard Thesaurus Wirtschaft (STW)  
<http://zbw.eu/stw/versions/8.10/download/> oder Thesaurus Sozialwissenschaften  
<http://www.gesis.org/unser-angebot/recherchieren/thesauri-und-klassifikationen/thesaurus-sozialwissenschaften/>



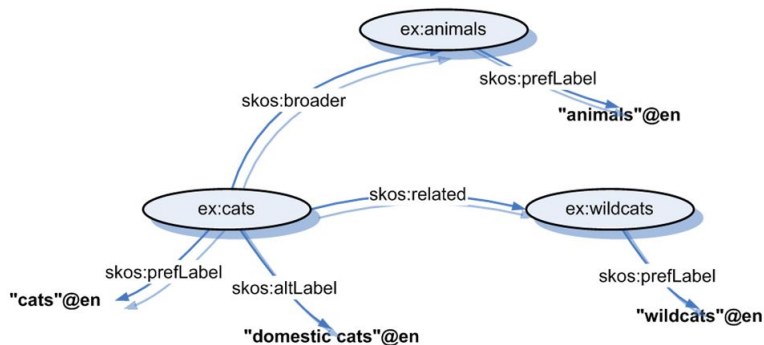
# Linked Open Data

## Die Prinzipien

- Vorangetrieben durch den „Erfinder“ des World Wide Web Tim Berners-Lee
- *Linked Data* als erster Schritt zum *Semantic Web*
  - Use URIs as names for things (and concepts)
  - Use HTTP URIs so that people can look up those names
  - When someone looks up an URI, provide useful information, using standards like RDF, SPARQL
  - Include links to other URIs, so they can discover more things
- HTTP-URIs identifizieren nicht mehr nur Webressourcen (HTML-Seiten, Dateien) sondern dienen auch zur Bezugnahmen auf Personen, Orte, Konzepte uvm.
- URIs sind öffentliche Identifikatoren und jeder/jede kann sie verwenden



# SKOS – Simple Knowledge Organization System



# SKOS – Simple Knowledge Organization System

## Vokabular

- Basic Structure
  - skos:Concept
- Lexical Labelling
  - skos:prefLabel, skos:altLabel, skos:hiddenLabel
- Symbolic Labelling
  - skos:prefSymbol, skos:altSymbol
- Documentation
  - skos:definition skos:note
- Semantic Relations
  - skos:broader, skos:narrower, skos:related

